

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-192392

(43)Date of publication of application : 28.07.1995

(51)Int.Cl.

G11B 20/02

G10L 3/02

(21)Application number : 06-109873

(71)Applicant : SANYO ELECTRIC CO LTD

(22)Date of filing : 24.05.1994

(72)Inventor : TANAKA KOJI
SUGISHITA SHOZO
IIDA MASAYUKI
MIYATAKE MASANORI

(30)Priority

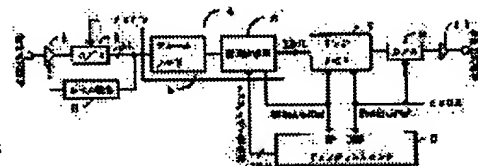
Priority number : 05255040
05312580Priority date : 18.09.1993
17.11.1993Priority country : JP
JP

(54) SPEAKING SPEED CONVERSION DEVICE

(57)Abstract:

PURPOSE: To obtain a speaking speed conversion device which lightens process load, reduces the deviation between an image and a speech, and has not so large increase in the capacity of a memory for storing speech signal.

CONSTITUTION: The speaking speed converting means 6 of the speaking speed conversion device is equipped with section decision means 11, 12, and 13 which decide whether an input speech corresponding to a necessary number of speech signals inputted to the frame memory 5 is in a voiced section or voiceless section, and signal processing means 21 to 26 which perform a compressing/ expanding process or deleting process for said necessary number of speech signals according to the outputs of the section decision means 11, 12, and 13 and the output of a storage quantity calculating means 9.



(11)特許出願公開番号

特開平7-192392

(43)公開日 平成7年(1995)7月28日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 1 B 20/02	G	9294-5D		
G 1 0 L 3/02	A			

審査請求 未請求 請求項の数15 O.L (全 25 頁)

(21)出願番号 特願平6-109873

(22)出願日 平成6年(1994)5月24日

(31)優先權主張番号 特願平5-255040

(32)優先日 平5(1993)9月18日

(33)優先権主張国 日本 (JP)

(31)優先権主張番号 特願平5-312580

(32)優先日 平5(1993)11月17日

(33)優先権主張国 日本 (J P)

(71) 出願人 000001889

三洋電機株式会社

大阪府守口市京阪本通2丁目5番5号

(72)発明者 田中 浩司

大阪府守口市京阪本通2丁目5番5号 三

洋電機株式会社内

(72) 發明者 杉下 正蔵

大阪府守口市京阪本通2丁目5番5号 三

洋電機株式会社内

(72)発明者 飯田 正幸

大阪府守口市京阪本通2丁目5番5号 三

洋電機株式会社内

(74) 代理人 弁理士 香山 秀幸

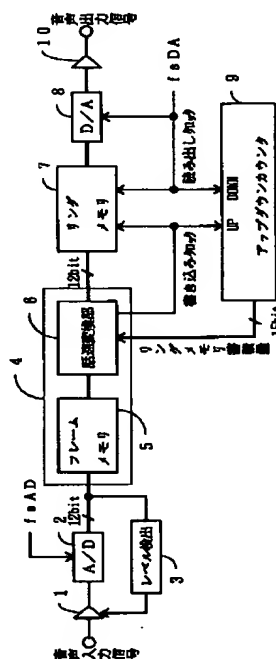
最終頁に続く

(54) 【発明の名称】 話速変換装置

(57) 【要約】

【目的】 この発明は、処理負荷を低減できるとともに、映像と音声のズレを小さくでき、しかも音声信号を蓄積するためのメモリの容量も膨大とならない話速変換装置を提供することを目的とする。

【構成】 話速変換装置において、話速変換処理手段6は、フレームメモリ5に入力された所要数の音声信号に対応する入力音声、音声区間か無音区間かを判別する区間判別手段11、12、13、ならびに、区間判別手段11、12、13の出力および蓄積量算出手段9の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段21～26を備えている。



【特許請求の範囲】

【請求項1】入力音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう話速変換装置。

【請求項2】入力音声信号を話速変換処理する話速変換処理手段、

話速変換処理手段の出力が書き込まれるリングメモリ、およびリングメモリからデータを一定速度で読み出す手段を備え、

話速変換処理手段は、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なう手段を備えている話速変換装置。

【請求項3】入力されるアナログ音声信号を設定された再生速度倍率に応じたサンプリング周波数でサンプリングするA/D変換手段、

A/D変換手段から出力された音声信号が入力されるフレームメモリ、

フレームメモリに所要数の音声信号が入力されるごと

に、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、

話速変換処理手段の出力が書き込まれるリングメモリ、リングメモリから一定速度でデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、

話速変換処理手段は、

フレームメモリに入力された所要数の音声信号に対応する入力音声、音声区間か無音区間かを判別する区間判別手段、ならびに、

区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段、

を備えている話速変換装置。

【請求項4】入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度で書き込まれるフレームメモリ、

フレームメモリに所要数の音声信号が入力されるごと

に、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、

話速変換処理手段の出力が書き込まれるリングメモリ、1倍速再生時のフレームメモリへの書き込み速度と等しい周波数の読み出し信号に基づいて、リングメモリからデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、

話速変換処理手段は、

フレームメモリに入力された所要数の音声信号に対応する入力音声、音声区間か無音区間かを判別する区間判別手段、ならびに、

区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理を行なう信号処理手段、

を備えている話速変換装置。

【請求項5】上記信号処理手段は、区間判別手段の出力および蓄積量算出手段の出力に基づいて、

(1)入力音声、音声区間でありかつリングメモリがオーバーフロー直前状態でない第1モード、

(2)入力音声、音声区間でありかつリングメモリがオーバーフロー直前状態である第2モード、

(3)入力音声、無音区間でありかつ無音区間の継続長が所定の無音削除開始点判別値未満であり、かつリングメモリがオーバーフロー直前状態でない第3モード、

(4)入力音声、無音区間でありかつ無音区間の継続長が所定の無音削除開始点判別値未満であり、かつリングメモリがオーバーフロー直前状態である第4モード、

(5)入力音声、無音区間でありかつ無音区間の継続長が所定の無音削除開始点判別値以上であり、かつリングメモリがアンダーフロー直前状態でない第5モード、および

(6)入力音声、無音区間でありかつ無音区間の継続長が所定の無音削除開始点判別値以上であり、かつリングメモリがアンダーフロー直前状態である第6モード、

のうちのいずれのモードであるかを判別するモード判別手段、第1モードまたは第3モードと判別されたときに、設定再生速度倍率を n として、音声信号に対して、圧縮率 $1/n$ より大きな圧縮率で圧縮伸長処理を行なう第1処理手段、

第2モードまたは第4モードと判別されたときに、リングメモリの蓄積量がアンダーフロー直前状態となるまで音声信号を削除する第2処理手段、

第5モードと判別されたときに、無音区間の音声信号を削除する第3処理手段、ならびに、

第6モードと判別されたときに、設定再生速度倍率を n として、圧縮率 $1/n \pm \alpha$ （ただし、 α は0以上で1以下の値）で圧縮伸長処理を行なう第4処理手段、

を備えていることを特徴とする請求項3および4のいずれかに記載の話速変換装置。

【請求項6】上記区間判別手段は、

フレームメモリに入力された所要数の音声信号のパワー平均値を算出する手段、および算出されたパワー平均値と所与のしきい値とに基づいて、入力音声、音声区間か無音区間かを判別する判別手段、

を備えている請求項3、4および5のいずれかに記載の話速変換装置。

【請求項7】上記区間判別手段は、

フレームメモリに入力された所要数の音声信号のパワー累積値を算出する手段、および算出されたパワー累積値

と所与のしきい値とに基づいて、入力音声の音声区間か無音区間かを判別する判別手段、
を備えている請求項3、4および5のいずれかに記載の
話速変換装置。

【請求項8】上記区間判別手段は、
フレームメモリに入力された所要数の音声信号の振幅平
均値を算出する手段、および算出された振幅平均値と所
与のしきい値とに基づいて、入力音声の音声区間か無音
区間かを判別する判別手段、
を備えている請求項3、4および5のいずれかに記載の
話速変換装置。

【請求項9】上記区間判別手段は、
フレームメモリに入力された所要数の音声信号の振幅累
積値を算出する手段、および算出された振幅累積値と所
与のしきい値とに基づいて、入力音声の音声区間か無音
区間かを判別する判別手段、
を備えている請求項3、4および5のいずれかに記載の
話速変換装置。

【請求項10】上記区間判別手段は、
フレームメモリに入力された所要数の音声信号の周期性
を検出する検出手段、および検出された周期に基づい
て、入力音声の音声区間か無音区間かを判別する判別手
段、
を備えている請求項3、4および5のいずれかに記載の
話速変換装置。

【請求項11】上記区間判別手段は、
フレームメモリに入力された所要数の音声信号の所定の
1または複数の周波数帯域に対するパワースペクトルを
算出する算出手段、および算出されたパワースペクトル
と所与のしきい値とに基づいて、入力音声の音声区間か
無音区間かを判別する判別手段、
を備えている請求項3、4および5のいずれかに記載の
話速変換装置。

【請求項12】上記しきい値が、上記リングメモリの蓄
積量に応じて調整される請求項6、7、8、9および1
1のいずれかに記載の話速変換装置。

【請求項13】上記第1処理手段は、ピッチ周期単位ま
たはピッチ周期の整数倍単位で圧縮伸長処理を行なうも
のである請求項5記載の話速変換装置。

【請求項14】上記第1処理手段は、固定フレーム長単
位で圧縮伸長処理を行なうものである請求項5記載の話
速変換装置。

【請求項15】上記無音削除開始点判別値が、上記リン
グメモリの蓄積量に応じて調整される請求項5に記載の
話速変換装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は、音声信号の話速を変
える話速変換装置に関し、例えば、映像を伴うレーザデ
ィスク、VTR等の音声の早聞きまたは遅聞きを行なう

音声再生装置、聴覚障害者に放送される音声信号をゆっ
くりした聞きやすい音声に変換する聴覚補助装置、ネイ
ティブスピードで話された英語音声をゆっくりした聞き
やすい音声に変換する英語学習器等に利用される話速変
換装置に関する。

【0002】

【従来の技術】話速を変換する従来の技術として、アナ
ログ方式の時間軸伸長圧縮技術がある。しかしながら、
アナログ方式の時間軸伸長圧縮技術を用いた話速変換方
法では、単純な音声波形の間引きまたは音声波形の繰り
返し挿入が行なわれているだけなので、音声のつながめ
が不連続になるため、音質が悪くなるという問題があ
る。

【0003】良好な音質が得られる音声の時間軸伸長圧
縮技術として、デジタル信号処理によって、音声のピ
ッチ周期を検出し、検出したピッチ周期単位またはピッ
チ周期の整数倍単位でピッチ部の間引きまたは挿入を行
なう技術がある。しかしながら、このデジタル方式の
時間軸伸長圧縮技術を用いた話速変換方法では、音声信
号における無音区間および音声区間にかかわらず、一律
の圧縮伸長率で音声信号の圧縮または伸長が行なわれて
いるので、VTRの倍速再生時、英語学習器の英語音声
再生時等においては音声区間の再生速度が速くなりす
ぎ、音声の聞き取れなくなる場合があるという問題があ
る。

【0004】

【発明が解決しようとする課題】上記問題を解決するた
めに、音声信号の無音区間と音声区間とを識別し、無音
区間を削除し、音声区間をピッチ周期単位で伸長する話
速変換方法が既に開発されている（参考文献A（以下、
第1従来方式という）：信学技法SP92-56、HC
92-33（1992-09） タイトル「話速変換に
伴う時間伸長を吸収するための一方法」 社団法人 電
子情報通信学会発行、参考文献B（以下、第2従来方式
という）：信学技法SP92-150（1993-0
3） タイトル「難聴者による話速変換方式の評価」
社団法人 電子情報通信学会発行）。この方法によれ
ば、音声区間の再生速度を遅くでき、音声の聞きやす
くなる。しかしながら、この方法では、次のような問題が
ある。

【0005】第1従来方式では、処理負荷が大きいた
め、高速演算が必要となり、消費電力が大きくなる。第
2従来方式では、映像と音声のズレが大きくなりすぎ内
容把握が困難となるとともに、音声信号を蓄積するた
めのメモリの容量が膨大となりコストがかかる。

【0006】この発明は、処理負荷を低減できるととも
に、映像と音声のズレを小さくでき、しかも音声信号を
蓄積するためのメモリの容量も膨大とならない話速変換
装置を提供することを目的とする。

【0007】

【課題を解決するための手段】この発明による第1の話速変換装置は、入力音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理を行なうことを特徴とする。

【0008】この発明による第2の話速変換装置は、入力音声信号を話速変換処理する話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、およびリングメモリからデータを一定速度で読み出す手段を備え、話速変換処理手段は、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に

【0009】この発明による第3の話速変換装置は、入力されるアナログ音声信号を設定された再生速度倍率に応じたサンプリング周波数でサンプリングするA/D変換手段、A/D変換手段から出力された音声信号が入力されるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、1倍速再生時のサンプリング周波数と等しい周波数の読み出し信号に基づいて、リングメモリからデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声

【0010】この発明による第4の話速変換装置は、入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度で書き込まれるフレームメモリ、フレームメモリに所要数の音声信号が入力されるごとに、それらの音声信号に対して話速変換処理を行なう話速変換処理手段、話速変換処理手段の出力が書き込まれるリングメモリ、リングメモリから一定速度でデータを読み出す読出手段、およびリングメモリの書き込み信号と読み出し信号とに基づいて、リングメモリの蓄積量を算出する蓄積量算出手段を備えており、話速変換処理手段は、フレームメモリに入力された所要数の音声信号に対応する入力音声

【0011】上記リングメモリとは、リング構造(ring structure)を有するメモリをいう。リング構造とは、連鎖リストの最後の項目のポインタが先頭の項目をさすよ

うにつながれたものをいう。

【0012】上記信号処理手段としては、たとえば、区間判別手段の出力および蓄積量算出手段の出力に基づいて、(1)入力音声

【0013】上記第1処理手段としては、ポインター移動量制御による重複加算法(Pointer Interval Control Overlap and Add: P I C O L A)、T D H S (Time Domain Harmonic Scaling)法等のように、ピッチ周期単位またはピッチ周期の整数倍単位で圧縮伸長処理を行なうものまたは固定フレーム長単位で圧縮伸長処理を行なうもの等が用いられる。

【0014】上記無音削除開始点判別値を、上記リングメモリの蓄積量に応じて調整するようにしてもよい。

【0015】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号のパワー平均値を算出する手段、および算出されたパワー平均値と所与のしきい値とに基づいて、入力音声

【0016】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号のパワー累積値を算出する手段、および算出されたパワー累積値と

所与のしきい値とに基づいて、入力音声の音声区間か無音区間かを判別する判別手段を備えているものが用いられる。上記しきい値を上記リングメモリの蓄積量に応じて調整するようにしてもよい。

【0017】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号の振幅平均値を算出する手段、および算出された振幅平均値と所与のしきい値とに基づいて、入力音声の音声区間か無音区間かを判別する判別手段を備えているものが用いられる。上記しきい値を上記リングメモリの蓄積量に応じて調整するようにしてもよい。

【0018】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号の振幅累積値を算出する手段、および算出された振幅累積値と所与のしきい値とに基づいて、入力音声の音声区間か無音区間かを判別する判別手段を備えているものが用いられる。上記しきい値を上記リングメモリの蓄積量に応じて調整するようにしてもよい。

【0019】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号の周期性を検出する検出手段、および検出された周期に基づいて、入力音声の音声区間か無音区間かを判別する判別手段を備えているものが用いられる。

【0020】上記区間判別手段としては、たとえば、フレームメモリに入力された所要数の音声信号の所定の1または複数の周波数帯域に対するパワースペクトルを算出する算出手段、および算出されたパワースペクトルと所与のしきい値とに基づいて、入力音声の音声区間か無音区間かを判別する判別手段を備えているものが用いられる。上記しきい値を上記リングメモリの蓄積量に応じて調整するようにしてもよい。

【0021】

【作用】この発明による第1の話速変換装置では、入力音声信号が音声区間であるか無音区間であるかに応じて、入力音声信号に対して圧縮伸長処理または削除処理が行われる。

【0022】この発明による第2の話速変換装置では、入力音声信号は話速変換処理手段によって、話速変換処理される。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、一定速度で読み出される。話速変換処理手段においては、入力音声信号が音声区間であるか無音区間であるかおよびリングメモリの蓄積量に応じて、入力音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【0023】この発明による第3の話速変換装置では、入力されるアナログ音声信号は、A/D変換手段により、設定された再生速度倍率に応じたサンプリング周波数でサンプリングされる。A/D変換手段から出力された音声信号は、フレームメモリに入力される。フレームメモリに所要数の音声信号が入力されるごとに、話速変

換処理手段により、それらの音声信号に対して話速変換処理が行なわれる。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、1倍速再生時のサンプリング周波数に等しい周波数の読み出し信号に基づいて読み出される。リングメモリの書き込み信号と読み出し信号とに基づいて、蓄積量算出手段によって、リングメモリの蓄積量が算出される。

【0024】話速変換処理手段においては、フレームメモリに入力された所要数の音声信号に対する入力音声の区間判別手段により、音声区間か無音区間かが判別される。そして、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【0025】この発明による第4の話速変換装置では、入力されるデジタル音声信号が、設定された再生速度倍率に応じた速度でフレームメモリに書き込まれる。フレームメモリに所要数の音声信号が入力されるごとに、話速変換処理手段により、それらの音声信号に対して話速変換処理が行なわれる。話速変換処理手段の出力は、リングメモリに書き込まれる。リングメモリに書き込まれたデータは、読み出し信号に基づいて一定速度で読み出される。リングメモリの書き込み信号と読み出し信号とに基づいて、蓄積量算出手段によって、リングメモリの蓄積量が算出される。

【0026】話速変換処理手段においては、フレームメモリに入力された所要数の音声信号に対する入力音声の区間判別手段により、音声区間か無音区間かが判別される。そして、区間判別手段の出力および蓄積量算出手段の出力に応じて、上記所要数の音声信号に対して圧縮伸長処理または削除処理が行なわれる。

【実施例】以下、図面を参照して、この発明をVTRに適用した場合の実施例について説明する。

【0027】図1は、話速変換装置の全体的な構成を示している。

【0028】入力音声信号は、ALCアンプ1で増幅された後、A/D変換部2に送られ、例えば12ビットのデジタル信号に変換される。A/D変換部2の標準サンプリング周波数は、たとえば8KHzである。2倍速再生時には、A/D変換部2のサンプリング周波数 f_s は、16KHzとなる。

【0029】A/D変換部2の出力は、DSP(Digital Signal Processor)4に送られるとともにレベル検出部3にも送られる。レベル検出部3は、A/D変換部2でA/D変換されたデータが変換レンジの最大値となったときに、ALC(automatic level control)信号をALCアンプ1に出力する。これにより、ALCアンプ1のアンプ利得が制御され、A/D変換部2の入力信号が最大レンジを越えないようにされる。つまり、VTRの再生テープ速度が変化するとALCアンプ1の入力信号

レベルも変化する。そこで、レベル検出部3の出力に基づいて、アンプ利得を自動調整することにより、A/D変換部2の入力信号が最大レンジを越えないようにしている。

【0030】DSP4は、2フレーム分の音声信号を記憶できる容量のフレームメモリ5およびフレームメモリ5に記憶された音声信号に対してフレーム単位で話速変換処理を行なう話速変換部6とを備えている。1フレームは、ここでは、200個のサンプリングデータから構成されるものとする。

【0031】フレームメモリ5内の前半領域および後半領域のうち、一方の領域に記憶された1フレーム分の音声信号に対して話速変換部6により処理が行なわれると同時に、他方の領域にA/D変換部2からの信号が蓄積される。そして、この他方の領域に1フレーム分の信号が蓄積されると、今度はその領域内のデータに対して話速変換部6により処理が行なわれると同時に、既に処理が行なわれたデータが記憶されていた上記一方の領域にA/D変換部2からの信号が蓄積される。

【0032】話速変換部6から出力されたデータは、書き込みクロックに基づいてリングメモリ7に書き込まれる。リングメモリ7に書き込まれたデータは、読み出しクロックに基づいて、読み出される。リングメモリ7から読み出された信号は、D/A変換部8によってアナログ信号に変換された後、アンプ10で増幅され、音声出力信号として出力される。

【0033】D/A変換部8のサンプリング周波数 f_{sDA} は、8KHzである。また、リングメモリ7の読み出しクロックの周波数も8KHzである。リングメモリ7としては、 $21845 \times 12 \text{ bit}$ のもの、すなわち、21845ワードのものが用いられている。したがって、リングメモリ7にデータを蓄積できる最大時間（入力信号に対する出力時間の最大遅延時間）は、 $21845 \times 1 / 8000 = 2.73$ 秒となる。

【0034】リングメモリ7に対する書き込みクロックは、アップダウンカウンタ9のアップカウント用入力端子（UP）に入力する。リングメモリ7に対する読み出しクロックは、アップダウンカウンタ9のダウンカウント用入力端子（DOWN）に入力する。アップダウンカウンタ9は、入力された書き込みクロックの総数と入力された読み出しクロックの総数との差（リングメモリ7の蓄積量）をカウントし、そのカウント値を15bitのデジタル信号として出力する。アップダウンカウンタ9の出力は、話速変換部6に送られる。

【0035】図2は、話速変換部6の詳細な構成を示している。

【0036】フレームメモリ5から読み出された音声信号は、パワー計算部11に送られ、1フレーム分の音声信号の平均パワー値Pが算出される。この平均パワー値Pは、サンプリングされた1フレーム内の各音声信号の

振幅を i_0, i_1, \dots, i_{N-1} （ただし、 $N=200$ ）とすると、次の数式1によって求められる。

【0037】

【数1】

$$P = \frac{1}{N} \times \sum_{k=0}^{N-1} (i_k)^2$$

10

【0038】パワー計算部11で求められた平均パワー値Pは、比較部12に送られる。比較部12には、しきい値メモリ13からしきい値 T_h が送られており、平均パワー値Pがしきい値 T_h 以上（ $P \geq T_h$ ）か、平均パワー値Pがしきい値 T_h より小さいか（ $P < T_h$ ）が判別される。比較部12からは、平均パワー値Pがしきい値 T_h 以上（ $P \geq T_h$ ）のときには現フレームが音声区間であることを示す信号が、平均パワー値Pがしきい値 T_h より小さいときには現フレームが無音区間であることを示す信号が、それぞれ出力される。

20

【0039】しきい値 T_h としては、A/D変換部2の量子化ビット数が12bitのときには、たとえば、 2^{12} に設定される。なお、次のようにして、しきい値 T_h を変更するようにしてもよい。すなわち、図2に点線で示すように、パワー定常状態検出およびしきい値更新部14を設ける。パワー定常状態検出およびしきい値更新部14は、パワー計算部11からの平均パワー値Pが、所定フレーム数（例えば、40フレーム）にわたって一定であったか否かを判別し、一定であったときには（定常状態）、そのときの平均パワー値Pの2倍の値をしきい値メモリ13に書き込み、しきい値 T_h を更新させる。ただし、更新されるしきい値の最大値は、所定値、たとえば 2^{14} に制限される。このようにすることにより、定常的に発生している雑音を無音区間として取り扱うことができるようになる。

30

【0040】また、入力信号の音声区間と無音区間とを、次の数式2で示す各フレームの音声信号のパワー累積値Paと所与のしきい値とに基づいて判別するようにしてもよい。

【0041】

【数2】

$$Pa = \sum_{k=0}^{N-1} (i_k)^2$$

40

【0042】比較部12の出力は、条件分岐部15に送られる。条件分岐部15には、リングメモリ蓄積量状態判別部16の出力が入力している。また、条件分岐部15には、パワー計算部11を介してフレームメモリ5からの、音声信号が送られている。さらに、条件分岐部15には、ポーズ継続長設定メモリ17が接続されている。ポーズ継続長設定メモリ17には、無音区間の削除開始点を決定するためのポーズ継続長 T_{del} （無音削除開始点判別値）が設定されている。

50

【0043】リングメモリ蓄積量状態判別部16は、アップダウンカウンタ9から送られてきた蓄積量に基づいて、リングメモリ7の状態がオーバーフロー直前状態になったこと、およびリングメモリ7の状態がアンダーフロー直前状態になったことを検出する。

【0044】つまり、オーバーフロー検出用データメモリ18にはオーバーフロー検出用データTmaxが、アンダーフロー検出用データメモリ19にはアンダーフロー検出用データTminが、それぞれ記憶されている。オーバーフロー検出用データTmaxは、例えば、リングメモリ7の総ワード数(TOTAL)21845より200小さい値21645に設定されている。アンダーフロー検出用データTminは、例えば、200に設定されている。

【0045】そして、アップダウンカウンタ9から送られてきた蓄積量がオーバーフロー検出用データTmax以上になると、リングメモリ蓄積量状態判別部16からオーバーフロー直前検出信号が出力される。また、アップダウンカウンタ9から送られてきた蓄積量がアンダーフロー検出用データTmin以下になると、リングメモリ蓄積量状態判別部16からアンダーフロー直前検出信号が出力される。条件分岐部15は、オーバーフロー直前検出信号が入力されているときにはリングメモリ7がオーバーフロー直前状態であると判別し、アンダーフロー直前検出信号が入力されているときにはリングメモリ7がアンダーフロー直前状態であると判別する。

【0046】条件分岐部15は、比較部12から送られてくる音声区間または無音区間の判別信号と、リングメモリ蓄積量状態判別部16から送られてくるリングメモリ状態に関する検出信号と、ポーズ継続長設定メモリ17に設定されているポーズ継続長Tdelとに基づいて、以下の6つのケースに場合分けを行なう。そして、それに応じて、マルチプレクサ20を制御して、音声信号を所定の処理部に送る。

【0047】(1)第1ケース(case1)
入力信号が音声区間であり、かつリングメモリ7がオーバーフロー直前状態ではないと判別されたときには、第1ケースとなる。

【0048】この場合には、音声信号は、マルチプレクサ20を介して、ピッチ圧縮伸長手段23に送られる。ピッチ圧縮伸長手段23は、バリエブルスピーチコントロール(VSC)を行なうものであり、再生速度倍率をnとすると、入力信号に対して、圧縮率1/nより大きな圧縮率で伸長圧縮処理を行なう。ここで用いられる伸長圧縮法としては、例えば、ポインター移動量制御による重複加算法(Pointer Interval Control Overlap and Add: PICO LA)、TDHS(TimeDomain Harmonic Scaling)法等がある。ピッチ伸長圧縮手段23で伸長圧縮処理が行なわれた信号は、デマルチプレクサ27を介してリングメモリ7に送られ、書き込みクロッ

クにしたがって、リングメモリ7に書き込まれる。

【0049】VTRの2倍速再生時においては、A/D変換部2のサンプリング周波数fsADは16KHzであり、D/A変換部8のサンプリング周波数fsDAは8KHzである。このため、音程は元に戻されて出力される。

【0050】従来の一般的な時間軸伸長圧縮においては、VTRの2倍速再生時には圧縮率1/2で、圧縮される。言い換えれば、2ピッチ周期が1ピッチ周期に間引かれる。このため、出力音声は標準音声速度の2倍速となる。つまり、2倍速再生の通常再生では、出力音声は標準音声速度の2倍速となる。ただし、音程は元のままとする。

【0051】これに対し、図2の話速変換部6に設けられた上記ピッチ伸長圧縮手段23では、圧縮率が1/2より大きな値に設定される。ここでは、圧縮率が2/3に設定されているとする。言い換えれば、3ピッチ周期が2ピッチ周期に間引かれる。このため、出力音声は、標準音声速度の3/2倍速となる。この場合も音程は、元のままだである。このように、圧縮率2/3で圧縮された場合には、圧縮率1/2の場合に比べて、2/3-1/2=1/6だけ、信号が伸長されることになる。この伸長分が、リングメモリ7の蓄積量となる。

【0052】PICO LAを用いて、入力信号を圧縮率2/3で圧縮する方法について、図3を用いて簡単に説明する。まず、入力信号からピッチ周期が抽出される。抽出されたピッチ周期をTpとする。波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)がつけられて、波形A'が作成される。波形Bに対しては0から1に向かう重み(重み関数K2)がつけられて、波形B'が作成される。

【0053】そして、それらの波形A'およびB'が加え合わされ、長さTpの波形A'*B'が作成される。これらの重みは、波形A'*B'の前後の接続点での連続性を保つためにつけられている。つぎに、ポインターが、圧縮率に基づいて決まる長さである3Tp分だけ移動され、同様な操作が行なわれる。これにより、3つの波形A、B、Cから2つの波形A'*B'およびCが得られる。このようにして、3ピッチ周期分の信号が、2ピッチ周期分の信号に圧縮される。

【0054】ピッチ伸長圧縮手段23による伸長圧縮法としては、図17(a)、(b)に示すように、ピッチ抽出をすることなく、所定長の固定フレーム長Ts単位で伸長圧縮処理を行うようにしてもよい。固定フレーム長Tsは、たとえば入力データの200個分の長さに設定される。図17の例では、3Tsを2Tsにする例を示している。

【0055】図17(a)の方法では、固定フレーム長Tsの波形A、B、Cのうち、波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)がつけられ

て、波形A"が作成される。波形Bに対しては0から1に向かう重み(重み関数K2)がつけられて、波形B"が作成される。

【0056】そして、それらの波形A"およびB"が加え合わされ、長さTsの波形A"*B"が作成される。これらの重みは、波形A"*B"の前後の接続点での連続性を保つためにつけられている。そして、次の波形Cに対しては、そのまま出力される。これにより、3つの波形A、B、Cから2つの波形A"*B"およびCが得られる。このようにして、3Ts分の信号が、2Ts分の信号に圧縮される。

【0057】図17(b)の方法では、固定フレーム長Tsの波形A~Cのうちの波形Aには先頭からたとえば20個のデータに0から1へ直線的に向かう重み(重み関数K3)をつけて波形A"を得る。波形Bには181個目~200個目までの入力データに1から0へ直線的に向かう重み(重み関数K4)をつけて波形B"を得る。そして、波形Cを削除する。次の3つの波形D~Fに対しても、同様な処理が行われる。このようにして、3つの波形A~C(またはD~F)からなる信号は、2つの波形A"およびB"(またはD"およびE")からなる信号に圧縮される。つまり、3Ts分の信号が、2Ts分の信号に圧縮される。

【0058】上記固定フレーム長単位での伸長圧縮処理を用いた場合には、ピッチ周期ごとの伸長圧縮処理を用いた場合に比べて、音質は低下するが、処理量は軽減される。

【0059】なお、この話速変換装置が英語学習器に適用されている場合には(1倍速再生時)、A/D変換部2のサンプリング周波数fsADは8KHzであり、D/A変換部8のサンプリング周波数fsDAは8KHzである。この場合には、ピッチ圧縮伸長手段23で、たとえば、2ピッチ周期が3ピッチ周期になるように、圧縮率3/2で音声信号が伸長される。つまり、音声区間が1.5倍に伸長される。したがって、この場合には、1倍速再生の通常再生時に対して、 $3/2 - 1 = 1/2$ だけ信号が伸長されることになり、この伸長分がリングメモリ7の蓄積量となる。

【0060】(2)第2ケース(case2)
入力信号が音声区間であり、かつリングメモリ7がオーバーフロー直前状態であると判別されたときには、第2ケースとなる。

【0061】この場合には、音声信号はマルチプレクサ20を介して、入力信号削除部21に送られ、音声信号が削除される。具体的には、アップダウンカウンタ9のカウンタ値が、アンダーフロー検出用データTmin以下になるまで、すなわちリングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込み動作が停止される。

【0062】リングメモリ7がアンダーフロー直前状態

になると、200個以下の個数、例えば100個の消音信号(値"0"の信号)が消音挿入部22から出力され、この消音信号がデマルチプレクサ27を介してリングメモリ7に送られて書き込まれる。このように、消音信号をリングメモリ7へ書き込んでいるのは、音声削除によって音声信号の繋ぎ目にクリック音が発生するのを防止するためである。

【0063】(3)第3ケース(case3)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel未満であり、かつリングメモリ7がオーバーフロー直前状態ではないと判別されたときには、第3ケースとなる。

【0064】この場合は、上記第1ケースの場合と同じ処理が行なわれる。ただし、第3ケースに該当する場合には、再生速度倍率をnとすると、 $1/n$ の圧縮率で伸長圧縮処理を行ってもよい。つまり、第3ケースに該当する場合には、 $1/n$ 以上の圧縮率で伸長圧縮処理が行われる。

【0065】(4)第4ケース(case4)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel未満であり、かつリングメモリ7がオーバーフロー直前状態であると判別されたときには、第4ケースとなる。

【0066】この場合は、上記第2ケースの場合と同じ処理が行なわれる。

【0067】(5)第5ケース(case5)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長Tdel以上であり、かつリングメモリ7がアンダーフロー直前状態ではないと判別されたときには、第5ケースとなる。

【0068】この場合には、音声信号はマルチプレクサ20を介して、入力信号削除部25に送られ、音声信号が削除される。具体的には、リングメモリ7への書き込み動作が停止される。ただし、音声区間のスタート部分(無声区間)が欠落するのを防止したり、音声の削除によって繋ぎ目にクリック音が発生したりするのを防止するために、波形合成挿入部26によって波形合成挿入処理が行なわれる。

【0069】波形合成挿入部26による波形合成挿入処理について、図4(a)、(b)を用いて説明する。図4(a)による方法では、波形合成挿入部26は、第1メモリ31および第2メモリ32を備えている。入力信号削除部25による入力信号削除処理の開始時においては、削除開始点から、1フレーム長以下の所定長さTs、例えば1フレーム分の入力信号が、第1メモリ31にアドレス順に順次記憶される。次に、第1メモリ31のアドレスが大きくなるにしたがって1から0に直線的に変化する関数K1が、第1メモリ31の内容Aに乗算される。そして、その乗算結果A'が、再度第1メモリ31に書き込まれる。

【0070】また、入力信号削除部25による入力信号削除区間の終了点直前の所定長さ T_s 分の入力信号が、第2メモリ32にアドレス順に順次記憶される。次に、第2メモリ32のアドレスが大きくなるほど、0から1に直線的に変化する関数 K_2 が、第2メモリ32の内容 B に乗算される。そして、その乗算結果 B' が、再度第2メモリ32に書き込まれる。この後、第1メモリ31の内容 A' と、第2メモリ32の内容 B' とが加え合わされて、所定長さ T_s のデータ $A' * B'$ が得られる。そして、得られた所定長さ T_s 分のデータ $A' * B'$ がデマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。

【0071】図4(b)による方法では、削除開始点から、1フレーム長以下の所定長さ T_s 、例えば1フレーム分の入力信号が、第1メモリ31にアドレス順に順次記憶される。次に、後端に1から0に直線的に変化するスロープがついた関数 K_3 が、第1メモリ31の内容 A に乗算される。そして、その乗算結果 A' が、再度第1メモリ31に書き込まれる。

【0072】また、入力信号削除部25による入力信号削除区間の終了点直前の所定長さ T_s 分の入力信号が、第2メモリ32にアドレス順に順次記憶される。次に、前端に0から1に直線的に変化するスロープがついた関数 K_4 が、第2メモリ32の内容 B に乗算される。そして、その乗算結果 B' が、再度第2メモリ32に書き込まれる。この後、第1メモリ31の内容 A' と、第2メモリ32の内容 B' とが繋ぎ合わされて、 $2T_s$ 分のデータ $A' + B'$ が得られる。そして、得られた $2T_s$ 分のデータ $A' + B'$ がデマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。図4(b)では、 T_s が、1フレーム分の長さである例を示したが、1フレームの半分の長さのデータを T_s としてもよい。

【0073】なお、入力信号削除部25による無音区間の音声信号の削除処理が繰り返し行なわれている場合に、リングメモリ7がアンダーフロー直前状態になることがある。この場合には、リングメモリ7がアンダーフロー直前状態になったときから、所定長さ T_s 分の入力信号が第2メモリ32に記憶される。そして、第1メモリ31に記憶されているデータと、第2メモリ32に記憶されているデータにもとづいて、上記と同様な波形合成挿入処理が行なわれる。

【0074】(6)第6ケース(case 6)

入力信号が無音区間であり、かつ無音区間の継続長が設定されたポーズ継続長 T_{del} 以上であり、かつリングメモリ7がアンダーフロー直前状態であると判別されたときには、第6ケースとなる。

【0075】この場合は、入力信号は、マルチプレクサ20を介して間引き処理部24に送られる。間引き処理部24では、VTRの再生速度倍率を n として、圧縮率

が $1/n$ となるように間引き処理が行なわれる。たとえば、2倍速再生時には入力信号に対して圧縮率 $1/2$ で間引きが行なわれ、3倍速再生時には入力信号に対して圧縮率 $1/3$ で間引きが行なわれる。1倍速再生時には、入力信号がそのまま出力される。

【0076】 $1/n$ 間引き処理部24による間引き処理としては、次のような方法が用いられる。ここでは、2倍速再生時を例にとって説明する。

【0077】上述したPICOLAまたはTDHSを用いた時間軸圧縮法を用い、入力信号のピッチを抽出し、ピッチデータ部分を圧縮率が $1/2$ となるように、間引く。

【0078】また、図5(a)～(c)に示すように、ピッチ抽出をすることなく、所定時間 T_s ごとに波形を間引くようにしてもよい。

【0079】図5(a)の方法では、波形A～Dのうち、波形Bおよび波形Dが間引かれ、波形A、Cからなる信号が得られる。

【0080】図5(b)の方法では、波形A～Dのうち、波形Bと波形Dが間引かれている。また、波形Aには、前端に0から1に上昇するスロープ(関数 K_4)が、後端に1から0に下降するスロープ(関数 K_3)がついた関数が乗算されて、波形 A' が作成される。また、波形Cには、前端に0から1に上昇するスロープ(関数 K_4)が、後端に1から0に下降するスロープ(関数 K_3)がついた関数が乗算されて、波形 C' が作成される。このようにして、4つの波形A～Dからなる信号は、2つの波形 A' および C' からなる信号に圧縮される。

【0081】図5(c)の方法では、波形Aに対しては、1から0へ直線的に向かう重み(重み関数 K_1)がつけられて、波形 A' が作成される。波形Bに対しては0から1に向かう重み(重み関数 K_2)がつけられて、波形 B' が作成される。そして、それらの波形 A' および B' が加え合わされ、長さ T_s の波形 $A' * B'$ が作成される。

【0082】同様に、波形Cに対しては、1から0へ直線的に向かう重み(関数 K_1)がつけられて、波形 C' が作成される。波形Dに対しては0から1に向かう重み(関数 K_2)がつけられて、波形 D' が作成される。そして、それらの波形 C' および D' が加え合わされ、長さ T_s の波形 $C' * D'$ が作成される。このようにして、4つの波形A～Dからなる信号は、2つの波形 $A' * B'$ および $C' * D'$ からなる信号に圧縮される。

【0083】上述のように、第6ケースに該当する場合には、VTRの再生倍率を n として、圧縮率 $1/n$ で間引き処理が行われているが、次のようにして圧縮率を制御するようにしてもよい。

【0084】圧縮率 $1/n$ で間引き処理が行われている場合、D/A変換器8のサンプリング周波数 f_{sDA} と

A/D変換器2のサンプリング周波数 f_{sAD} との比 f_{sDA}/f_{sAD} が、圧縮率 $1/n$ と等しい場合には、リングメモリ7の蓄積量は、変化しない。しかしながら、圧縮率 $1/n$ の演算精度、サンプリング周波数 f_{sAD} と f_{sDA} のクロック精度によっては、 f_{sDA}/f_{sAD} が圧縮率 $1/n$ と等しくならないことが起こりうる。

【0085】 f_{sDA}/f_{sAD} が圧縮率 $1/n$ より大きくなったとき($f_{sDA}/f_{sAD} > 1/n$)には、 $f_{sDA}/f_{sAD} = 1/a$ ($a > 0$)として、 $\{(1/a) - (1/n)\}$ だけ、圧縮率が小さくなり、間引きの度合いが大きくなり、リングメモリ7の蓄積量が減少していき、リングメモリ7の蓄積量がアンダーフローするおそれがある。

【0086】一方、 f_{sDA}/f_{sAD} が圧縮率 $1/n$ より小さくなったとき($f_{sDA}/f_{sAD} < 1/n$)には、 $f_{sDA}/f_{sAD} = 1/a$ ($a > 0$)として、 $\{(1/n) - (1/a)\}$ だけ、圧縮率が大きくなり、間引きの度合いが小さくなり、リングメモリ7の蓄積量が増加していく。

【0087】したがって、間引き処理を行う場合には、リングメモリ7の蓄積量を確認して、次のように圧縮率を制御する。 $f_{sDA}/f_{sAD} = 1/a$ ($a > 0$)として、 $(1/n) - \alpha < 1/a < (1/n) + \alpha$ の条件を満たす α を選定する。ただし、 α は、0以上で1以下の値であり、例えば0.001~0.1の範囲の値である。

【0088】 f_{sDA}/f_{sAD} が圧縮率 $1/n$ より大きくなったとき、すなわち、リングメモリ7の蓄積量が減少していく場合には、圧縮率を $1/n$ から $\{(1/n) + \alpha\}$ にする。つまり、圧縮率を大きくし、リングメモリ7の蓄積量を増加させるようにする。

【0089】 f_{sDA}/f_{sAD} が圧縮率 $1/n$ より小さくなったとき、すなわち、リングメモリ7の蓄積量が増加していく場合には、圧縮率を $1/n$ から $\{(1/n) - \alpha\}$ にする。つまり、圧縮率を小さくし、リングメモリ7の蓄積量を減少させるようにする。

【0090】上記では、リングメモリ7の蓄積量に基づいて、圧縮率を変化させているが、間引き処理が行われる場合に、フレーム毎に圧縮率を $\{(1/n) - \alpha\}$ または $\{(1/n) + \alpha\}$ に、交互に変化させるようにしてもよい。

【0091】図6および図7は、話速変換部6による処理手順を示している。

【0092】以下、VTRの2倍速再生時の場合の話速変換部6による処理について、説明する。

【0093】(1) 再生開始時の処理

再生が開始されて、パワー計算部11によって最初のフレームの平均パワー値Pが算出されると(ステップ1)、算出された平均パワー値Pがしきい値Th以上か

否かが比較部12の出力に基づいて判別される(ステップ2)。

【0094】入力音声信号が無音区間から開始した場合、最初のフレームにおいては、平均パワー値Pはしきい値Thより小さくなり、ステップ1に進む。そして、無音区間の継続長(無音区間が継続するフレーム数)が算出され、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tdel以上か否かが判別される(ステップ12)。このポーズ継続長Tdelは、たとえば、フレーム数にして4フレーム分の長さに設定されている。

【0095】最初のフレームに対する処理においては、無音区間の継続長がポーズ継続長Tdel未満であるので、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がアンダーフロー直前状態か否かが判別される(ステップ13、14)。

【0096】最初のフレームに対する処理においては、リングメモリ7は、アンダーフロー直前状態になっているので、フレームデータが間引き処理部24によって圧縮率 $1/2$ で間引かれ(ステップ28)、間引き処理後の圧縮データがリングメモリ7に書き込まれる。この後、ステップ1に戻る。

【0097】(2) 第1ケースとなる処理の説明
ステップ2で、平均パワー値Pがしきい値Th以上であると判別されたときには、今回のフレームが音声区間であると判断され、ステップ3に進む。ステップ3では、前フレームが削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。前フレームが削除区間でない場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。前フレームが削除区間である場合には、ステップ4および5の処理が行われた後、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。ステップ4および5の処理については、後述する。

【0098】ステップ7において、オーバーフロー直前状態ではないと判別された場合には、第1ケースとなり、ピッチ圧縮伸長手段23によって、今回のフレームデータが2/3の圧縮率で時間軸圧縮される(ステップ8)。圧縮データは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

【0099】(2) 第2ケースとなる処理の説明
ステップ2で、平均パワー値Pがしきい値Th以上であると判別されたときには、今回送られてきたフレームは音声区間であると判断され、ステップ3に進む。ステップ3では、前フレームが削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。前フレームが削除区間でない場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、

7)。前フレームが削除区間である場合には、ステップ4および5の処理が行なわれた後、リングメモリ7がオーバーフロー直前状態か否かが判別される(ステップ6、7)。ステップ4および5の処理については、後述する。

【0100】ステップ7において、オーバーフロー直前状態であると判別された場合には、第2ケースとなり、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除される(ステップ9)。つまり、リングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込みが停止される。

【0101】そして、リングメモリ7がアンダーフロー直前状態になると、消音挿入部22によって、200個以下の所定数の消音信号“0”がリングメモリ7に書き込まれる(ステップ10)。そして、ステップ1に戻る。

【0102】上記ステップ10の処理の代わりに、図9(a)または図9(b)に示すような処理を行なってもよい。図9(a)に示す方法について説明すると、ステップ7でオーバーフロー直前状態と判別されたときから、たとえば、200個の入力信号に対する波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形A'を得る。また、アンダーフロー直前から200個前までの200個の入力信号に対する波形Bに対しては0から1に向かう重み(重み関数K2)をつけて、波形B'を得る。

【0103】そして、得られた2つの波形A'およびB'を加え合わせて、200個分の長さの波形A'*B'を作成する。そして、この波形A'*B'に対する200個の信号をリングメモリ7に書き込む。なお、アンダーフロー直前から200個前の時点の検出は、アップダウンカウンタ9のカウント値に基づいて行なわれる。これにより、音声削除区間の前後の音声信号の緊ぎ目にクリック音が発生するのを、効果的に防止できる。

【0104】図9(b)に示す方法について説明すると、ステップ7でオーバーフロー直前状態と判別されたときから、たとえば、100個の入力信号に対する波形Aに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形A'を得る。また、アンダーフロー直前から100個前までの100個の入力信号に対する波形Bに対しては0から1に向かう重み(重み関数K2)をつけて、波形B'を得る。そして、得られた2つの波形A'およびB'を繋ぎ合わせた200個分の信号をリングメモリ7に書き込む。

【0105】上記ステップ9では、オーバーフロー直前状態であると判別された場合には、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除されているが、リングメモリ7に蓄積されているデータ

を、リングメモリ7がアンダーフロー直前状態になるように、削除するようにしてもよい。

【0106】具体的には、リングメモリ7の書込開始アドレスを、図18(a)に示すオーバーフロー直前状態の時のアドレス(C地点)から、図18(b)に示すようにリングメモリ7がアンダーフロー直前状態となるアドレス(A地点)までジャンプさせる。したがって、ステップ9の処理では、A地点からC地点までのアドレスに蓄積されていたデータが削除されることになる。この後、図18(c)に示すように、ステップ10によって消音信号が書き込まれた後、入力データが書き込まれていく。

【0107】ステップ9において、上記のように、リングメモリ7に蓄積されているデータをリングメモリ7がアンダーフロー直前状態になるように削除した場合、ステップ10で消音信号をリングメモリ7に書き込む代わりに図19(a)または図19(b)のような処理を行ってもよい。

【0108】今、リングメモリ7の書込開始アドレスが、図18(a)に示すオーバーフロー直前状態の時のアドレス(C地点)から、図18(b)に示すようにリングメモリ7がアンダーフロー直前状態となるアドレス(A地点)までジャンプしたとする。このA地点から所定数、例えば200先のアドレス(図19(a)のB地点)までに蓄積されているデータSに対しては、図19(a)に示すように、1から0へ直線的に向かう重み(重み関数K1)をつけて波形S'を得る。また、それ以後にリングメモリ7に書き込まれる200個分の入力データ(波形T)に対しては、図19(a)に示すように、0から1に向かう重み(重み関数K2)をつけて、波形T'を得る。

【0109】そして、得られた2つの波形S'およびT'を加え合わせて、200個分の長さの波形S'*T'を作成する。そして、この波形S'*T'に対する200個の信号をA地点からリングメモリ7に書き込む。これにより、蓄積データ削除区間の前後の音声信号の緊ぎ目にクリック音が発生するのを、効果的に防止できる。

【0110】図19(b)に示す方法について説明すると、図18(b)のA地点から所定数、例えば100個先のアドレス(図19(b)のB地点)までに蓄積されているデータSに対しては、1から0へ直線的に向かう重み(重み関数K1)をつけて波形S'を得る。また、それ以後にリングメモリ7に書き込まれる100個分の入力データ(波形T)に対しては、0から1に向かう重み(重み関数K2)をつけて、波形T'を得る。そして、得られた2つの波形S'およびT'を繋ぎ合わせた200個分の信号をA地点からリングメモリ7に書き込む。

【0111】(3)第3ケースとなる処理の説明

21

ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され（ステップ11）、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される（ステップ12）。そして、無音区間の継続長がポーズ継続長Tde1未満であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される（ステップ13、14）。

【0112】リングメモリ7がアンダーフロー直前状態になっていないときには、リングメモリ蓄積量状態判別部16の出力に基づいて、オーバーフロー直前状態か否かが判別される（ステップ6、7）。オーバーフロー直前状態でない場合には、第3ケースとなり、ピッチ圧縮伸長手段23によって、今回のフレームデータが2/3の圧縮率で時間軸圧縮される（ステップ8）。圧縮データは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

【0113】（4）第4ケースとなる処理の説明
ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され（ステップ11）、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される（ステップ12）。そして、無音区間の継続長がポーズ継続長Tde1未満であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される（ステップ13、14）。

【0114】リングメモリ7がアンダーフロー直前状態になっていないときには、リングメモリ蓄積量状態判別部16の出力に基づいて、オーバーフロー直前状態か否かが判別される（ステップ6、7）。オーバーフロー直前状態である場合には、第4ケースとなり、リングメモリ蓄積量状態判別部16からアンダーフロー検出信号が出力されるまで、入力信号削除部21によって入力信号が削除される（ステップ9）。つまり、リングメモリ7がアンダーフロー直前状態になるまで、リングメモリ7への書き込みが中断される。

【0115】そして、リングメモリ7がアンダーフロー直前状態になると、消音挿入部22によって、200個以下の所定数の消音信号"0"がリングメモリ7に書き込まれる（ステップ10）。そして、ステップ1に戻る。

【0116】（5）第5ケースとなる処理の説明
ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され（ステップ11）、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tde1以上か否かが判別される（ステップ12）。そして、無音区間の継続長がポーズ継続長Tde1以上であると

22

判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される（ステップ15、16）。

【0117】リングメモリ7がアンダーフロー直前状態でないときには、第5ケースとなり、今回のフレームが入力信号削除部25による削除区間であることを示す第1フラグF1がセットされる（ステップ17）。この第1フラグF1は、電源投入時の初期設定において、リセット（F1=0）されている。そして、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるか否かを示す第2フラグF2がリセットされているか否かが判別される（ステップ18）。

【0118】この第2フラグF2は、電源投入時の初期設定において、リセット（F2=0）されている。そして、入力信号削除部25による削除区間の最初のフレームに対する処理が終了したときにセット（F2=1）にされる。そして、入力信号削除部25による一連の削除区間に対する処理が終了したときにリセット（F2=0）される。

【0119】したがって、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるときには、第2フラグF2は、リセット（F2=0）されている。第2フラグF2がリセットされているときには、波形合成挿入部26によって第1メモリ31に今回のフレームデータが記憶される（ステップ19）。また、入力信号削除部25によって今回のフレームデータのリングメモリ7への書き込みが停止される（ステップ20）。つまり、今回のフレームデータが削除される。そして、第2フラグF2がセット（F2=1）された後（ステップ21）、ステップ1に戻る。

【0120】さらに、無音区間が続いている場合には、ステップ2、11、12、15を通してステップ16に移り、リングメモリ蓄積量状態判別部16の出力に基づいて、リングメモリ7がアンダーフロー直前状態か否かが判別される。

【0121】リングメモリ7がアンダーフロー直前状態でないときには、今回のフレームが入力信号削除部25による削除区間であることを示す第1フラグF1がセットされる（ステップ17）。そして、今回のフレームが入力信号削除部25による削除区間の最初のフレームであるか否かを示す第2フラグF2がリセットされているか否かが判別される（ステップ18）。

【0122】この場合には、第2フラグF2はセット（F2=1）されているので、今回のフレームが入力信号削除部25による削除区間の最初のフレームでないと判断される。この場合には、波形合成挿入部26によって第2メモリ32に今回のフレームデータが記憶される（ステップ22）。また、入力信号削除部25によって今回のフレームデータのリングメモリ7への書き込みが停止される（ステップ23）。そして、ステップ1に戻

10

20

30

40

50

る。

【0123】そして、さらに、無音区間が続きかつリングメモリ7がアンダーフロー直前状態となっていないときには、ステップ2、11、12、15、16、17、18、22および23の処理が繰り返される。つまり、第2メモリ32のフレームデータが更新されるとともに、フレームデータのリングメモリ7への書き込みが停止される。

【0124】この後、音声区間のフレームデータが入力されたときには、ステップ2において、平均パワー値Pがしきい値Th以上となるので、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1状態に基づいて判別される（ステップ3）。この場合には、第1フラグF1がセット（F1=1）されているので、前フレームが入力信号削除部25による削除区間であったと判別され、ステップ4に移る。ステップ4では、入力信号削除部25による削除処理が停止せしめられるとともに、波形合成挿入部26による波形合成挿入処理が行なわれる。

【0125】すなわち、図4（a）を用いて既に説明したように、第1メモリ31の内容に1から0に直線的に変化する関数が乗算され、第2メモリ32の内容に0から1に直線的に変化する関数が乗算され、これらの両乗算結果が加え合わされる。この加算結果（図4（a）のA' * B'に相当する。）が、デマルチプレクサ27を介して、リングメモリ7に送られ、リングメモリ7に書き込まれる。

【0126】この後、第1フラグF1および第2フラグF2がリセット（F1=F2=0）される（ステップ5）。そして、ステップ6に進む。

【0127】ところで、連続している無音区間に対して、上記のような入力信号削除部25による削除処理が繰り返し行なわれている場合において、リングメモリ7がアンダーフロー直前状態になることがある。この場合には、上記ステップ16でYESとなり、ステップ24に移る。ステップ24では、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1の状態に基づいて判別される。

【0128】この場合には、第1フラグF1がセット（F1=1）されているので、ステップ25に進み、第2メモリ32に今回のフレームデータが記憶される。そして、入力信号削除部25による削除処理が停止せしめられるとともに、波形合成挿入部26による波形合成挿入処理が行なわれる（ステップ26）。そして、第1フラグF1および第2フラグF2がリセット（F1=F2=0）された後（ステップ27）、ステップ1に進む。

【0129】上記ステップ26における波形合成挿入部26による波形合成挿入処理には、上記ステップ4で説明した波形合成挿入処理とほぼ同様であるが、第2メモリ32に記憶されているフレームデータが、リングメモ

リ7がアンダーフロー直前状態になった後のフレームデータである点が、上記ステップ4で説明した処理の場合と異なっている。

【0130】なお、上記ステップ25の処理を省略し、ステップ24でYESとなった場合に、第2メモリ32に今回のフレームデータを記憶させることなく、ステップ26に移るようにしてもよい。この場合には、ステップ26で行なわれる波形合成挿入処理においては、上記ステップ4で説明した波形合成挿入処理と同様に、第2メモリ32に記憶されているアンダーフロー直前状態より前のフレームデータ（前回のフレームデータ）が用いられる。

【0131】また、上記ステップ22の処理を省略するとともに上記ステップ3と上記ステップ4との間に、フレームデータを第2メモリ32に記憶させるステップを追加するようにしてもよい。この場合には、ステップ4においては、上記ステップ19において第1メモリ31に記録された内容と、上記ステップ3と上記ステップ4との間に追加されたステップにおいて第2メモリ32に記録された内容とに基づいて、波形合成挿入処理が行われる。

【0132】（6）第6ケースとなる処理の説明
ステップ2で平均パワー値Pがしきい値Thより小さいと判別されたときには、今回までの無音区間の継続長が算出され（ステップ11）、算出された継続長がポーズ継続長メモリ17に設定されているポーズ継続長Tdel以上か否かが判別される（ステップ12）。そして、無音区間の継続長がポーズ継続長Tdel以上であると判別された場合には、リングメモリ蓄積量状態判別部16の出力に基づいて、アンダーフロー直前状態か否かが判別される（ステップ15、16）。

【0133】リングメモリ7がアンダーフロー直前状態であるときには、前フレームが入力信号削除部25による削除区間であったか否かが、第1フラグF1の状態に基づいて判別される（ステップ24）。第1フラグF1がリセットされている場合（F1=0）、すなわち、前フレームが入力信号削除部25による削除区間でなかった場合には、第6ケースとなり、ステップ28に移る。ステップ28では、間引き処理部24によって、今回のフレームデータが圧縮率1/2で間引き処理が行なわれる。そして、間引き処理されたデータは、リングメモリ7に送られて書き込まれる。この後、ステップ1に戻る。

【0134】つまり、無音区間の継続長がポーズ継続長Tdel以上であっても、リングメモリ7がアンダーフロー直前状態であり、かつ前フレームが入力信号削除部25による削除区間でない場合には、フレームデータは削除されず、圧縮率1/2で間引き処理が行なわれた後、リングメモリ7に書き込まれる。

【0135】図7においては、ステップ12において、

10

20

30

40

50

無音区間の継続長が設定されたポーズ継続長 T_{del} より長いかが判別されているが、図8のステップ12 Aに示すように、無音区間の継続長 T が設定された第1基準長 T_1 未満 ($T < T_1$)、無音区間の継続長 T が設定された第1基準長 T_1 以上で設定された第2基準長 T_2 (ただし $T_1 < T_2$) 未満 ($T_1 \leq T < T_2$)、または無音区間の継続長 T が設定された第2基準長 T_2 以上 ($T \geq T_2$) を、判別するようにしてもよい。第1基準長としては、たとえば、4フレーム分の長さが、第2基準長としてはたとえば40フレーム分の長さが設定される。

【0136】そして、図8に示すように、各判別結果に応じて、次のようなステップに進むようにしてもよい。すなわち、無音区間の継続長 T が設定された第1基準長 T_1 未満 ($T < T_1$) である場合には、ステップ13に進む。無音区間の継続長 T が設定された第1基準長 T_1 以上で設定された第2基準長 T_2 ($T_1 < T_2$) 未満 ($T_1 \leq T < T_2$) であるときには、ステップ28に進んで1/n間引き処理による間引きを行なう。無音区間の継続長 T が設定された第2基準長 T_2 以上 ($T \geq T_2$) であるときには、ステップ15に進む。

【0137】図10は、2倍速再生時の入力信号と出力信号との関係を示し、特に無音区間の入力信号が削除される様子を示している。図11および図12は、リングメモリ7へのデータ書き込み開始点、リングメモリ7からのデータ読み出し開始点ならびに図10の各点A~Hにおけるリングメモリ7の状態を示している。

【0138】図10では、2倍速再生開始時においては、入力信号は無音区間となっており、かつリングメモリ7は空状態であるので(図11(a)参照)、フレームデータが間引き処理部24によって圧縮率1/2で間引かれた後、リングメモリ7に書き込まれていく。

【0139】そして、リングメモリ7の蓄積量 T_m がアンダーフロー検出用データ T_{min} に達すると、リングメモリ7からのデータの読み出しが開始される(図11(b)参照)。

【0140】そして、入力信号の音声区間aに対するフレームデータが送られてくると(A点)、ピッチ圧縮伸長手段23によって、フレームデータが圧縮率2/3で圧縮される。入力信号と出力信号との長さが一致する圧縮率1/2の圧縮を基準とすると、フレームデータが伸長される。この意味で、図10には、伸長処理と記載されている。そして、この圧縮データがリングメモリ7に書き込まれる。A点においては、図11(c)に示すように、蓄積量 $T_m A$ は、 T_{min} のままである。

【0141】入力信号の音声区間aに対する出力信号a1は、A点での蓄積量 $T_m A$ 分だけ遅れて読み出されていく。そして、入力信号の音声区間aが入力され終わった時点(B点)では、図11(d)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、

A点からB点までの音声区間aの圧縮データの、圧縮率1/2の圧縮に対する伸長分 $S_t B$ との和がリングメモリ7の蓄積量 $T_m B (= S_t B + T_{min})$ となる。したがって、入力信号の音声区間aに対する出力信号a1は、B点から $T_m B (= S_t B + T_{min})$ 分が経過した点で出力され終わる。

【0142】入力信号の音声区間aに続くポーズ継続長 T_{del} 未満の無音区間のフレームデータも、ピッチ圧縮伸長手段23によって圧縮率2/3で圧縮される。この無音区間に続いて音声区間bが入力されると、この音声区間bのフレームデータもピッチ圧縮伸長手段23によって圧縮率2/3で圧縮される。

【0143】そして、入力信号の音声区間bが入力され終わった時点(C点)では、図11(e)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、A点からC点までの入力信号に対応する圧縮データの、1/2圧縮に対する伸長分 $S_t C$ との和がリングメモリ7の蓄積量 $T_m C (= S_t C + T_{min})$ となる。したがって、入力信号の音声区間bに対する出力信号b1は、C点から $T_m C (= S_t C + T_{min})$ 分が経過した点で出力され終わる。

【0144】入力信号の音声区間bに続いて、ポーズ継続長 T_{del} 以上の長さの無音区間の信号が送られてきたときには、ポーズ継続長 T_{del} に達するまで(D点)はフレームデータが、ピッチ圧縮伸長手段23によって圧縮率2/3で圧縮される。

【0145】D点では、図11(f)に示すように、今回の圧縮区間の開始点であるA点での蓄積量 T_{min} と、A点からD点までの入力信号に対応する圧縮データの、1/2圧縮に対する伸長分 $S_t D$ との和がリングメモリ7の蓄積量 $T_m D (= S_t D + T_{min})$ となる。したがって、入力信号の音声区間bとD点との間の無音区間に対する出力信号は、D点から $T_m D (= S_t D + T_{min})$ 分が経過した点で出力され終わる。

【0146】ポーズ継続長 T_{del} 以降の無音区間のフレームデータは、リングメモリ7の蓄積量がアンダーフロー検出用データ T_{min} 以下になるまで、入力信号削除部25によって削除される。このポーズ削除部分の長さ $S_t d$ は、今回の圧縮区間の開始点であるA点からD点までの入力信号に対応する圧縮データの、1/2圧縮に対する伸長分 $S_t D$ と等しくなる。入力信号削除部25によって削除処理が行なわれた後においては、波形合成挿入部22によってクリック音防止のための合成波形が挿入されるが、図10には挿入された合成波形部分を省略してある。

【0147】入力信号が削除された区間の最終点(E点)においては、図12(g)に示すように、リングメモリ7の蓄積量 $T_m E$ は、アンダーフロー検出用データ T_{min} 以下となる。ここでは、蓄積量 $T_m E$ が、アンダーフロー検出用データ T_{min} に等しくなった例を示

10

20

30

40

50

している。

【0148】E点からの無音区間に対するフレームデータは、間引き処理部24によって、圧縮率 $1/2$ で間引かれた後、フレームメモリ7に書き込まれる。そして、音声区間cの信号が入力されると(F点)、この音声区間cのフレームデータがピッチ圧縮伸長手段23によって、圧縮率 $2/3$ で圧縮される。つまり、新たな圧縮区間が開始される。そして、圧縮データがリングメモリ7に書き込まれる。

【0149】F点では、図12(h)に示すように、リングメモリ7の蓄積量 T_{mF} は、E点のときと同じ T_{min} となっている。

【0150】入力信号の音声区間cに対する出力信号c1は、F点での蓄積量 T_{min} 分だけ遅れて出力されていく。入力信号の音声区間cに続くポーズ継続長 T_{del} 未満の無音区間(音声区間cからG点までの無音区間)のフレームデータも、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮される。

【0151】G点では、図12(i)に示すように、今回の圧縮区間の開始点であるF点での蓄積量 T_{min} と、F点からG点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tG} との和がリングメモリ7の蓄積量 T_{mG} ($=S_{tG}+T_{min}$)となる。したがって、入力信号の音声区間cからG点までの無音区間に対する出力信号は、G点から T_{mG} ($=S_{tG}+T_{min}$)分が経過した点で出力され終わる。

【0152】ポーズ継続長 T_{del} 以降の無音区間のフレームデータは、リングメモリ7の蓄積量がアンダーフロー検出用データ T_{min} になるまで、入力信号削除部25によって削除される。このポーズ削除部分の長さ S_{td} は、今回の圧縮区間の開始点であるF点からG点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tG} と等しくなる。

【0153】入力信号が削除された区間の最終点(H点)においては、図12(j)に示すように、リングメモリ7の蓄積量 T_{mH} は、アンダーフロー検出用データ T_{min} 以下となる。ここでは、蓄積量 T_{mH} が、アンダーフロー検出用データ T_{min} に等しくなった例を示している。

【0154】H点からの無音区間に対するフレームデータは、間引き処理部24によって、圧縮率 $1/2$ で間引かれた後、フレームメモリ7に書き込まれる。そして、音声区間dの信号が入力されると、この音声区間dのフレームデータがピッチ圧縮伸長手段23によって、圧縮率 $2/3$ で圧縮される。そして、伸長されたデータがリングメモリ7に書き込まれる。

【0155】図13は、2倍速再生時の入力信号と出力信号との関係を示し、特にオーバーフロー直前状態となったときに、入力信号が削除される様子を示している。

図14は、図13の各点S~Uにおけるリングメモリ7

の状態を示している。

【0156】ある時点からT点までの、音声区間a、b、c等と無音区間とを含む一連の入力信号に対するフレームデータが、ピッチ圧縮伸長手段23によって圧縮率 $2/3$ で圧縮され(圧縮率 $1/2$ の圧縮に対しては伸長され)ているとする。この場合には、リングメモリ7に伸長分が蓄積されていく。

【0157】音声区間bの入力開始点(S点)においては、図14(a)に示すように、当該1連の入力信号の圧縮処理の開始点での蓄積量 T_{min} と、上記圧縮処理の開始点からS点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tS} との和がリングメモリ7の蓄積量 T_{mS} ($=S_{tS}+T_{min}$)となる。したがって、音声区間bに対する出力信号b1は、S点から T_{mS} ($=S_{tS}+T_{min}$)分が経過した点で出力され始められる。

【0158】音声区間cの入力信号に対応する圧縮データがリングメモリ7に書き込まれた時点(T点)において、リングメモリ7がオーバーフロー直前状態になったとする。すなわち、T点において、リングメモリ7の蓄積量がオーバーフロー検出用データ T_{max} 以上になったとする。

【0159】T点においては、図14(b)に示すように、当該1連の入力信号に対する圧縮処理の開始点での蓄積量 T_{min} と、上記圧縮処理開始点からT点までの入力信号に対応する圧縮データの、 $1/2$ 圧縮に対する伸長分 S_{tT} との和がリングメモリ7の蓄積量 T_{mT} ($=S_{tT}+T_{min}$)となる。言い換えれば、リングメモリ7の全ワード数をTOTALとし、オーバーフロー検出用データを T_{max} とし、TOTALと T_{max} との差を D_{min} とすると、T点での蓄積量 T_{mT} は、 T_{max} に等しいので、 $TOTAL-D_{min}$ となる。

【0160】したがって、当該1連の入力信号に対する出力信号は、T点から蓄積量 T_{mT} ($=S_{tT}+T_{min}$)分遅れた時点で出力され終わる。

【0161】T点において、リングメモリ7がオーバーフロー直前状態になると、それ以後の入力信号に対しては、リングメモリ7がアンダーフロー直前状態になるまで、入力信号削除部21によって無条件に削除される。

入力信号削除部21によって削除処理が行なわれた後においては、消音挿入部22によって消音が挿入されるが、図13には挿入された消音部分を省略してある。リングメモリ7がオーバーフロー直前状態になった後(T点)、フレームデータが削除されていき、図14(c)に示すようにU点でリングメモリ7がアンダーフロー直前状態(蓄積量 $T_{mU}=T_{min}$)になったとする。この場合には、T点からU点までの4つの無音区間および3つの音声区間d、e、fからなる入力信号が削除される。したがって、T点からU点までの入力信号は、出力信号としては現れない。

10

20

30

40

50

29

【0162】U点の後に音声区間gの信号が入力されると、この音声区間に対するフレームデータは、ピッチ圧縮伸長手段23によって圧縮率2/3で圧縮され（圧縮率1/2の圧縮に対しては伸長され）た後、リングメモリ7に書き込まれていく。音声区間gに対する出力信号gは、U点でのリングメモリ7の蓄積量Tmin分だけ遅れて出力され始められる。

【0163】上記実施例では、入力信号の音声区間と無音区間とを、各フレームの平均パワー値Pに基づいて判別しているが、各フレームの平均振幅に基づいて判別するようにしてもよい。この場合には、図15に示すように、図2のパワー計算部11の代わりにフレーム単位で平均振幅値を計算する平均振幅計算部11Aが設けられ、しきい値メモリ13Aには、A/D変換部2の量子化ビット数が12bitのときには、たとえば、値 2^6 のしきい値が設定される。そして、平均振幅計算部11Aによって計算された平均振幅値と、しきい値メモリ13Aのしきい値とが、比較部12Aによって比較されることにより、音声区間か無音区間かが判別される。

【0164】つまり、平均振幅値がしきい値以上であれば音声区間と判別され、平均振幅値がしきい値未満であれば無音区間と判別される。フレーム単位の平均振幅値Wは、サンプリングされた1フレーム内の各音声信号の振幅を i_0, i_1, \dots, i_{N-1} （ただし、 $N=200$ ）とすると、次の数式3に基づいて算出される。

【0165】

【数3】

$$W = (1/N) \times \sum_{k=0}^{N-1} |i_k|$$

【0166】その他の処理については、図2の話速変換部6による処理と同じであるので、その説明を省略する。

【0167】なお、この場合においても、次のようにして、しきい値を変更するようにしてもよい。すなわち、図15に点線で示すように、平均振幅定常状態検出およびしきい値更新部14Aを設ける。平均振幅定常状態検出およびしきい値更新部14Aは、平均振幅計算部11Aからの平均振幅値Wが、所定フレーム数にわたって一定であったか否かを判別し、一定であったときには（定常状態）、そのときの平均振幅値Wの2倍の値をしきい値メモリ13Aに書き込み、しきい値を更新させる。ただし、更新されるしきい値の最大値は、所定値、たとえば 2^8 に制限される。

【0168】また、入力信号の音声区間と無音区間とを、次の数式4で示す各フレームの音声信号の振幅累積値Waと所与のしきい値とに基づいて判別するようにしてもよい。

【0169】

【数4】

30

 $k = N-1$

$$W_a = \sum_{k=0}^{N-1} |i_k|$$

 $k = 0$

【0170】また、入力信号の音声区間と無音区間とを、各フレームの信号の周期性を検出し、検出した周期が予め定められた音声信号のピッチ周期範囲内であれば、音声区間であると判別し、検出した周期が予め定められた音声信号のピッチ周期範囲外であれば無音区間であると判別するようにしてもよい。

10 【0171】この場合には、図16に示すように、図2のパワー計算部11の代わりに、自己相関法に基づいて、フレームごとの周期性を検出するピッチ周期検出部11Bが設けられ、しきい値メモリ13Bには、音声信号のピッチ周期範囲が設定される。そして、ピッチ周期検出部11Bで検出された周期と、しきい値メモリ13Bに設定された音声信号のピッチ周期範囲とが、比較部12Bによって比較される。

【0172】設定される音声信号のピッチ周期範囲は、再生速度により異なり、n倍速再生のときには、たとえば、 $66 \times n$ (Hz) ~ $320 \times n$ (Hz) の範囲に設定される。したがって、2倍速再生時には、音声信号のピッチ周期範囲は、 132 Hz ~ 640 Hz の範囲に設定される。その他の処理については、図2の話速変換部6による処理と同じであるので、その説明を省略する。

【0173】また、入力信号の音声区間と無音区間とを、各フレームの信号のパワースペクトルと、定常状態のパワースペクトルと比較することにより、判別するようにしてもよい。

30 【0174】この場合には、図20に示すように、図2のパワー計算部11の代わりに、フレームごとに所定の1または複数の周波数帯域に対するパワースペクトルを算出するパワースペクトル算出部11Cが設けられる。また、上記所定の1または複数の周波数帯域に対する定常状態のパワースペクトルがパワースペクトル記憶部13Cに記憶されている。

【0175】パワースペクトル記憶部13Cの内容は、パワースペクトル算出部11Cによって算出されたパワースペクトルの変化状態に基づいて、パワースペクトル定常状態検出部14Bが定常状態であることを検出したときには、検出された定常状態でのパワースペクトルに更新される。

【0176】入力信号がパワースペクトル算出部11Cに送られてくると、フレームごとに所定の1または複数の周波数帯域に対するパワースペクトルが算出される。そして、算出されたパワースペクトルと、パワースペクトル記憶部13Cに記憶されている定常状態のパワースペクトルとが比較部12Cによって比較される。

50 【0177】算出されたパワースペクトルが定常状態のパワースペクトルに対して、変動していれば、そのフレームは音声区間と判別される。逆に、算出されたパワー

スペクトルが定常状態のパワースペクトルに対して、変動していなければ、そのフレームは無音区間と判別される。

【0178】具体的には、パワースペクトル記憶部13Cには、上記所定の1または複数の周波数帯域に対する定常状態のパワースペクトルに基づいて、上記所定の1または複数の周波数帯域に対するしきい値が記憶される。そして、パワースペクトル記憶部13Cに記憶されている。パワースペクトル算出部11Cによって算出された上記所定の1または複数の周波数帯域に対するパワースペクトルと、パワースペクトル記憶部13Cに記憶されている対応するしきい値とが比較されることにより、入力信号が音声区間か無音区間かが判別される。

【0179】たとえば、定常状態のパワースペクトルが図21の(a)に示されているように、雑音のみのパワースペクトルであるとする。また、雑音が含まれていない音声のパワースペクトルが図21の(b)に示されているものとする。定常状態において、図21(a)のパワースペクトルで示される雑音が存在する場合に、図21(b)で示すパワースペクトルを持つ音声信号が入力すると、そのパワースペクトルは、図21(c)に示されるように、両者のパワースペクトルが合成されたものとなる。

【0180】したがって、たとえば、定常状態のパワースペクトルにおいてパワーが比較的小さい周波数帯域f_aおよびf_bに対するパワーは、音声区間のパワースペクトルにおいては大幅に増加する。つまり、定常状態のパワースペクトルにおいてパワーが比較的小さい1または複数の周波数帯域における定常状態のパワーと、入力信号のパワースペクトルの上記1または複数の周波数帯域におけるパワーとを比較することにより、入力信号が音声区間か無音区間かを判別することができる。

【0181】なお、定常状態の雑音が高い周波数帯域の雑音であると判明している場合には、雑音の影響の少な*

*い低い周波数帯域(例えば、4KHz以下の周波数帯域)に対するパワースペクトルを算出し、算出されたパワースペクトルが所定のしきい値以上か否かによって、入力信号が音声区間か無音区間かを判別することもできる。

【0182】また、各フレームのパワー平均値Pと、しきい値Thとを比較することにより、音声区間と無音区間とを判別する場合において、リングメモリ7の蓄積量に基づいて、しきい値Thを変化させるようにしてもよい。すなわち、リングメモリ7の蓄積量が少なくなるほど、言い換えれば、リングメモリ7の空領域が多くなるほど、音声区間の欠落部が少なくなるようにしきい値Thは小さくされる。これにより、出力音声が自然により近くなる。

【0183】つまり、図22に示すように、しきい値調整手段51を設ける。しきい値調整手段51は、リングメモリ蓄積量状態判別部16からリングメモリ7の蓄積量を得る。そして、得られたリングメモリ7の蓄積量を、D/A変換部8のサンプリング周波数で除することにより、蓄積時間Tmを算出する。そして、算出された蓄積時間Tmに基づいて、しきい値Thを決定し、しきい値メモリ13の内容を更新する。

【0184】より具体的に説明すると、リングメモリ蓄積量状態判別部16から得られたリングメモリ7の蓄積量がD/A変換部8のサンプリング周波数である8000で除されることにより、蓄積時間Tmが求められる。そして、予め作成された蓄積時間Tmに対するしきい値Thのデータに基づいて、蓄積時間Tmに対するしきい値Thが求められる。

【0185】次の表は、A/D変換部2の量子化ビット数が12bitである場合における蓄積時間Tmに対するしきい値Thのデータの一例を示している。

【0186】

【表1】

Tm	0.25～ 0.5 秒	0.5 ～ 0.75秒	0.75～ 1.0 秒	1.0 ～ 1.25秒	1.25～ 1.5 秒	1.5 ～ 1.75秒	1.75～ 2.5 秒	2.5 秒 以上
Th	2 ⁷	2 ⁸	2 ⁹	2 ¹⁰	2 ¹¹	2 ¹²	2 ¹³	2 ¹⁴

【0187】また、各フレームのパワー累積値Paとしきい値とを比較することにより、音声区間と無音区間とを判別する場合、各フレームの平均振幅値Wとしきい値とを比較することにより、音声区間と無音区間とを判別する場合、各フレームの振幅累積値Waとしきい値とを比較することにより、各フレームのパワースペクトルとしきい値とを比較することにより、音声区間と無音区間とを判別する場合にも、上記と同様に、リングメモリ7の蓄積量に基づいて、しきい値を変化させるようにしてもよい。

【0188】また、リングメモリ7の蓄積量に基づいて、無音区間の削除開始点を決定するためのポーズ継続長Tdelを変化させるようにしてもよい。すなわち、リングメモリ7の蓄積量が少なくなるほど、言い換えれば、リングメモリ7の空領域が多くなるほど、無音区間の削除部が少なくなるように、ポーズ継続長Tdelが長くされる。これにより、出力音声が自然により近くなる。

【0189】つまり、図22に示すように、ポーズ継続長調整手段52を設ける。ポーズ継続長調整手段52

は、リングメモリ蓄積量状態判別部16からリングメモリ7の蓄積量を得る。そして、得られたリングメモリ7の蓄積量を、D/A変換部8のサンプリング周波数で除することにより、蓄積時間 T_m を算出する。そして、算出された蓄積時間 T_m に基づいて、ポーズ継続長 T_{del} を決定し、ポーズ継続長設定メモリ17の内容を更新する。

【0190】より具体的に説明すると、リングメモリ蓄積量状態判別部16から得られたリングメモリ7の蓄積量がD/A変換部8のサンプリング周波数である 800×10

T_m	0.25～ 0.5 秒	0.5 ～ 0.75秒	0.75～ 1.0 秒	1.0 ～ 1.25秒	1.25～ 1.5 秒	1.5 ～ 1.75秒	1.75～ 2.5 秒	2.5 秒 以上
T_{del}	0.350 秒	0.300 秒	0.250 秒	0.200 秒	0.150 秒	0.100 秒	0.050 秒	0.025 秒

【0193】以上は、入力信号がアナログ信号の場合について説明したが、入力信号がデジタルデータである場合にもこの発明を適用することができる。たとえば、ICメモリ、磁気ディスク、デジタル通信回線等から、圧縮されたデジタル音声信号が送られてきた場合には、圧縮されたデジタル音声信号が伸長されてPCM音声信号に変換され、得られたPCM音声信号がバッファに一旦格納される。その後、設定された再生速度倍率に応じた速度で、PCM音声データがバッファから読み出されて、図1のフレームメモリ5に送られる。

【0194】

【発明の効果】この発明によれば、処理負荷を低減できるとともに、映像と音声のズレを小さくでき、しかも音声信号を蓄積するためのメモリの容量も膨大とならない話速変換装置が得られる。

【図面の簡単な説明】

【図1】話速変換装置の全体的な構成を示すブロック図である。

【図2】話速変換部の構成を示すブロック図である。

【図3】PICO Lを用いて、入力信号を圧縮率2/3で圧縮する方法を示す説明図である。

【図4】波形合成処理部による処理を説明するための説明図である。

【図5】間引き処理部によって行なわれる各種の間引き処理方法を説明するための説明図である。

【図6】話速変換部による処理手順を示すフローチャートである。

【図7】話速変換部による処理手順を示すフローチャートである。

【図8】話速変換部による処理手順の変形例を示し、図7に相当するフローチャートである。

【図9】図6のステップ10の処理と置き換え可能な処理を説明するための説明図である。

*0で除されることにより、蓄積時間 T_m が求められる。そして、予め作成された蓄積時間 T_m に対するポーズ継続長 T_{del} のデータに基づいて、蓄積時間 T_m に対するポーズ継続長 T_{del} が求められる。

【0191】次の表は、VTRの2倍速再生時における蓄積時間 T_m に対するポーズ継続長 T_{del} のデータの一例を示している。

【0192】

【表2】

【図10】2倍速再生時の入力信号と出力信号との関係を示し、特に無音区間の入力信号が削除される様子を示すタイムチャートである。

【図11】リングメモリ7へのデータ書き込み開始点、リングメモリ7からのデータ読み出し開始点ならびに図10の点A～Dにおけるリングメモリ7の状態を示す模式図である。

【図12】図10の点E～Hにおけるリングメモリ7の状態を示す模式図である。

【図13】2倍速再生時の入力信号と出力信号との関係を示し、特にオーバーフロー直前状態となったときに、入力信号が削除される様子を示すタイムチャートである。

【図14】図13の各点S～Uにおけるリングメモリ7の状態を示す模式図である。

【図15】音声区間と無音区間とを判別するための回路の変形例を示し、図2に相当するブロック図である。

【図16】音声区間と無音区間とを判別するための回路の他の変形例を示し、図2に相当するブロック図である。

【図17】固定フレーム単位で、入力信号を圧縮率2/3で圧縮する方法を示す説明図である。

【図18】図6のステップ9の処理と置き換え可能な処理を説明するための説明図である。

【図19】図6のステップ9の処理として図18の処理を採用した場合に、図6のステップ10の処理と置き換え可能な処理を説明するための説明図である。

【図20】音声区間と無音区間とを判別するための回路のさらに他の変形例を示し、図2に相当するブロック図である。

【図21】定常状態のパワースペクトル、雑音を含まない音声のパワースペクトルおよび音声区間のパワースペクトルを示すグラフである。

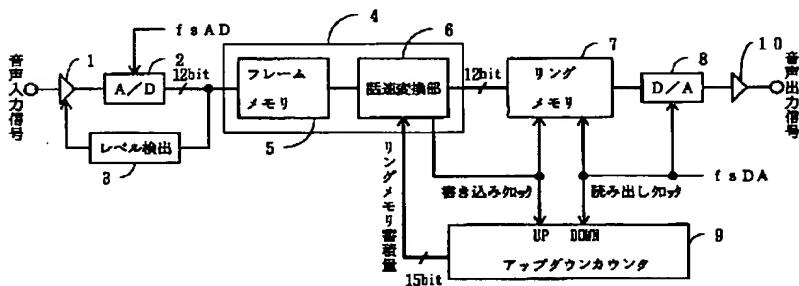
【図 22】しきい値調整手段およびポーズ継続長調整手段が付加された話速変換部を示すブロック図である。

【符号の説明】

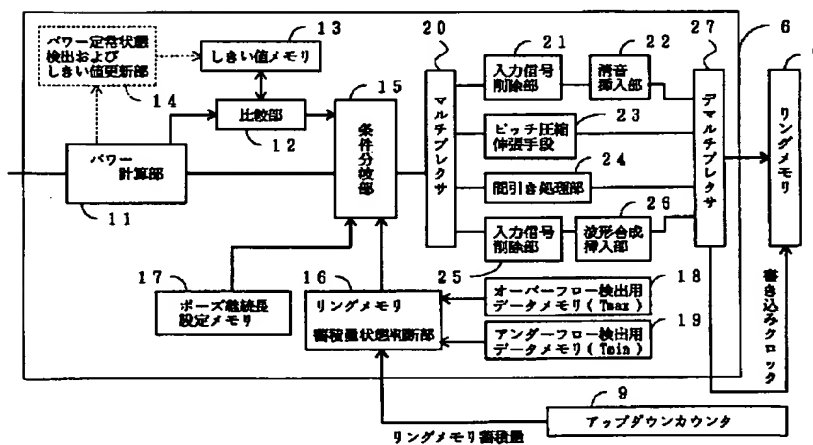
- 2 A/D変換部
4 DSP
5 フレームメモリ
6 話速変換部
7 リングメモリ
8 D/A変換部
9 アップダウンカウンタ
11 パワー計算部

- | | | | |
|----|---|-----------------------|----------------|
| 三 | * | 1 1 A | 平均振幅計算部 |
| | | 1 1 B | ピッチ周期検出部 |
| | | 1 1 C | パワースペクトル計算部 |
| | | 1 2、1 2 A、1 2 B、1 2 C | 比較部 |
| | | 1 5 | 条件分岐部 |
| | | 1 6 | リングメモリ蓄積量状態判別部 |
| | | 2 1、2 5 | 入力信号削除部 |
| | | 2 3 | ピッチ圧縮伸長手段 |
| | | 2 4 | 間引き処理部 |
| 10 | | 5 1 | しきい値調整手段 |
| * | | 5 2 | ポーズ継続長調整手段 |

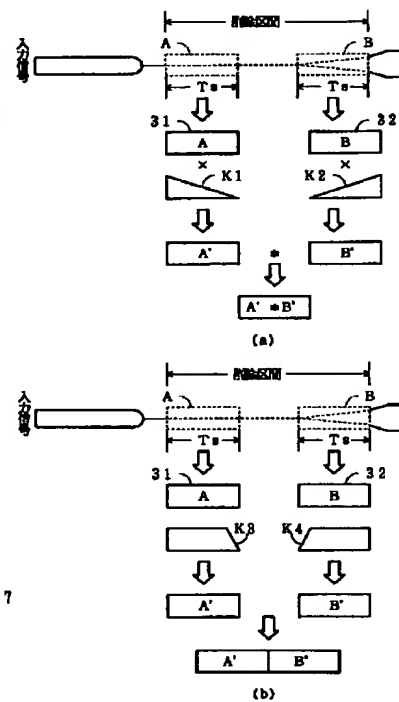
【図 1】



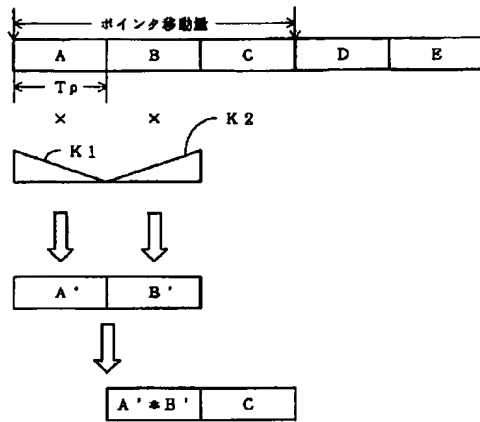
【図 2】



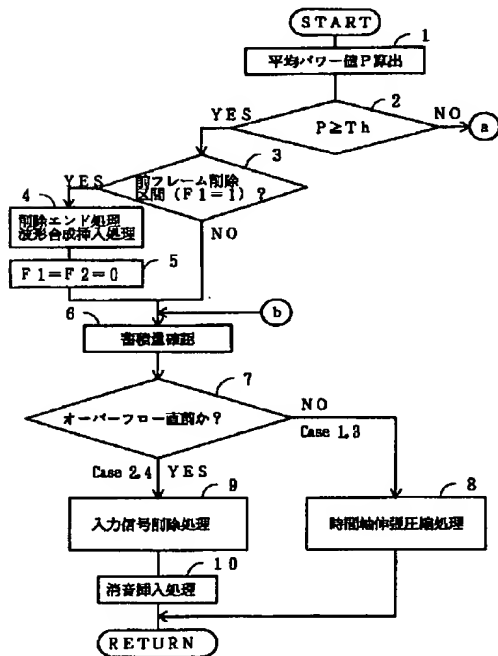
【図 4】



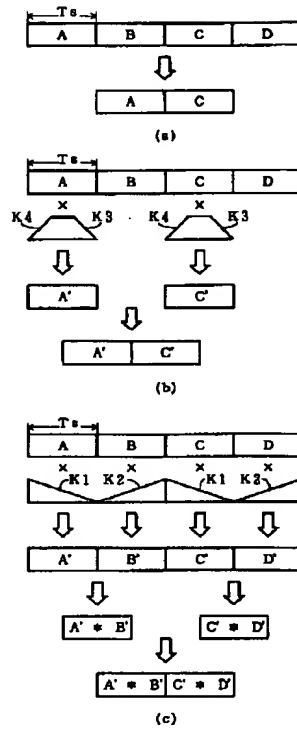
【図3】



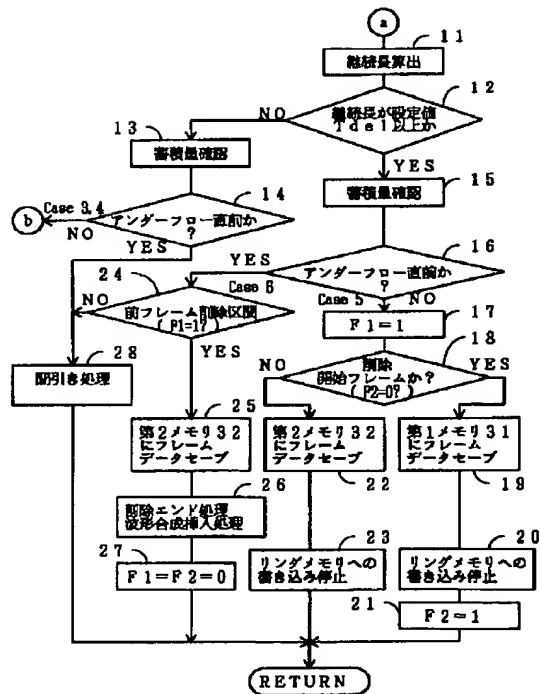
【図6】



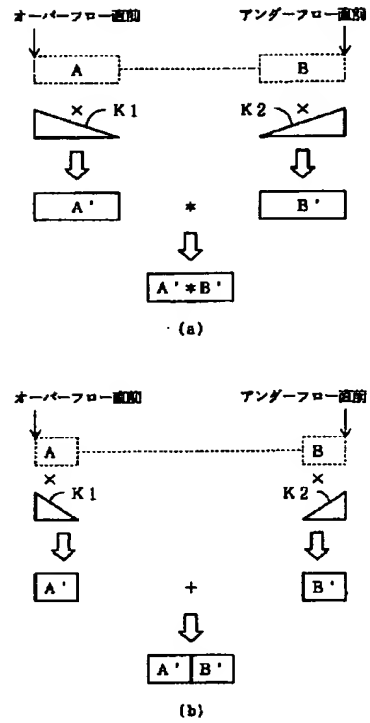
【図5】



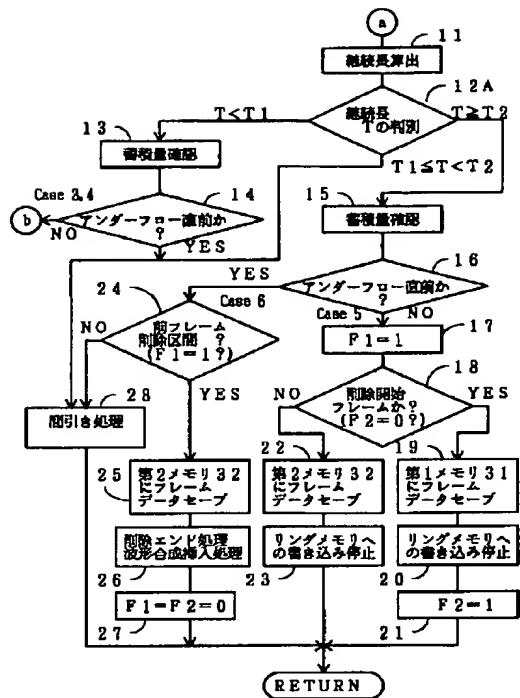
【図7】



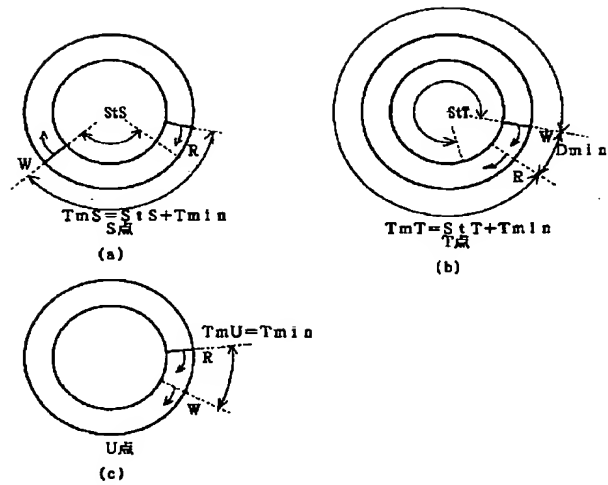
【図9】



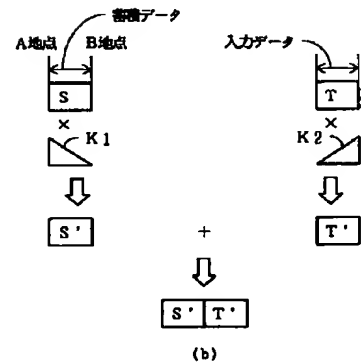
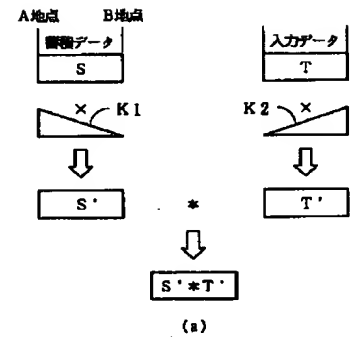
【図8】



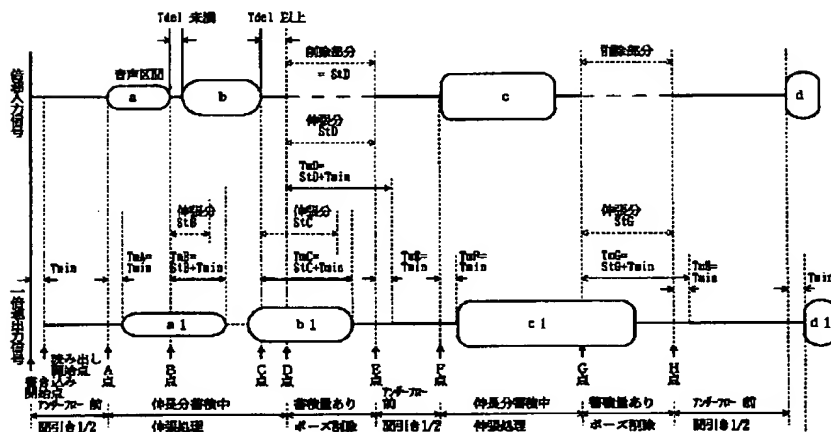
【図14】



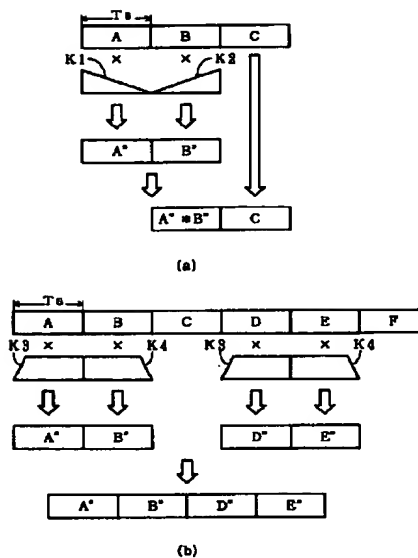
【図19】



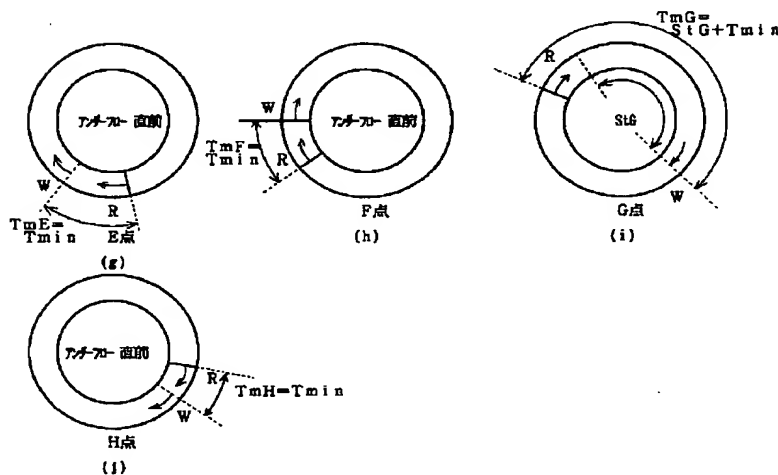
【図10】



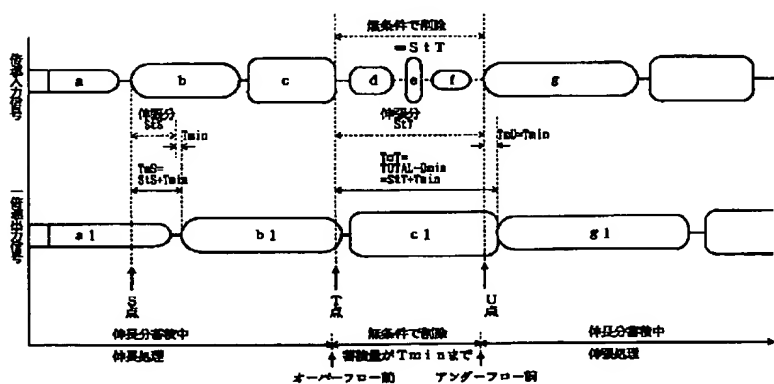
【圖 17】



【图 12】



【图 13】



【图 16】

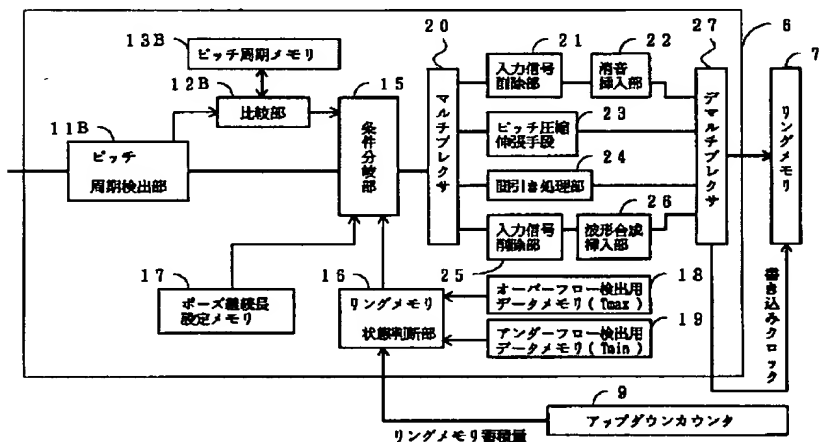
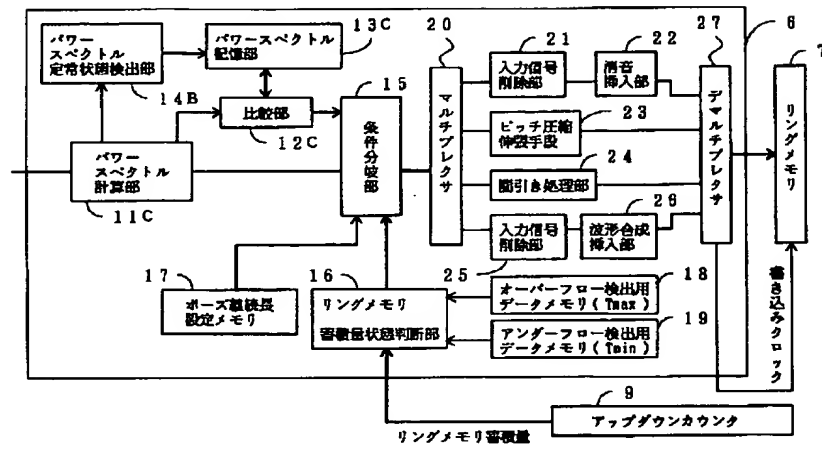
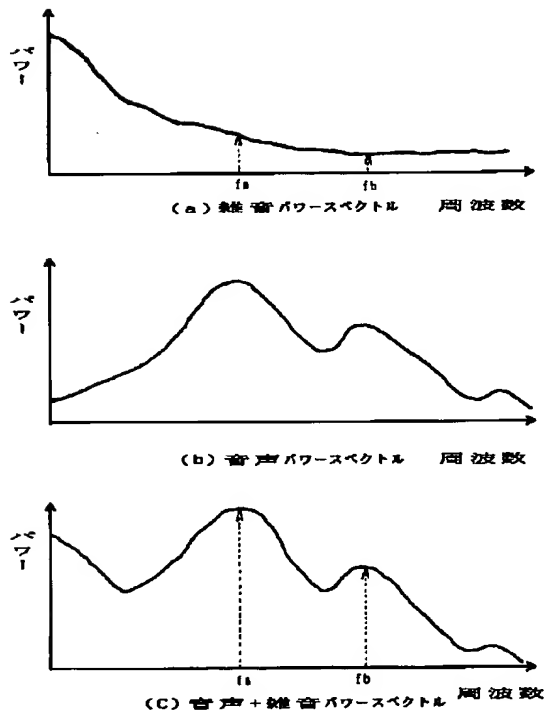


Figure 1 consists of three circular diagrams labeled (a), (b), and (c), each representing a roundabout with an inner circle and an outer ring. In all diagrams, 'R' is labeled at the top of the outer ring, representing the number of lanes. In (a), 'C地点' (Point C) is on the left and 'A地点' (Point A) is at the top, with arrows indicating vehicle flow 'W' entering from C and exiting towards A. Below (a) is the text 'オーバーフロー直前状態' (State just before overflow). In (b), 'C地点' is on the left and 'A地点' is on the right, with arrows indicating vehicle flow 'W' entering from C and exiting towards A. A shaded area on the right side of the ring is labeled '禁止区' (Prohibited zone). Below (b) is the text 'アンダーフロー直前状態' (State just before underflow). In (c), 'C地点' is on the left and 'A地点' is on the right, with arrows indicating vehicle flow 'W' entering from C and exiting towards A. Below (c) is the text '消去信号' (Clearance signal).

【図20】



【図21】



[illegible]

(72)発明者 宮武 正典
大阪府守口市京阪本通2丁目5番5号 三
洋電機株式会社内